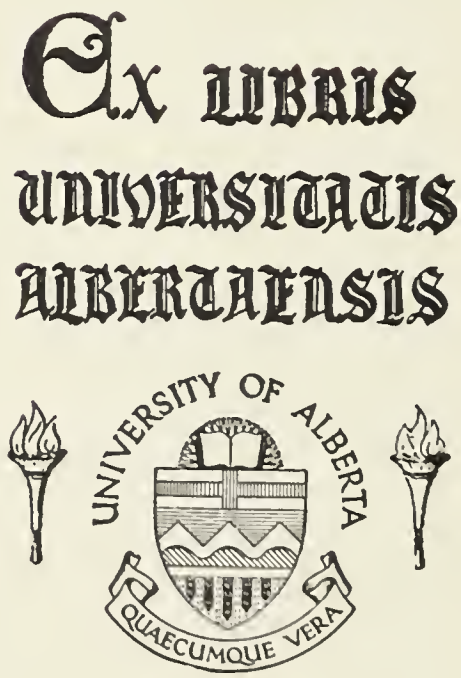


University of Alberta
Printing Department

THE UNIVERSITY OF ALBERTA

THE VALIDITY OF A SELF-ADMINISTERING WECHSLER SHORT FORM

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF ARTS

FACULTY OF ARTS AND SCIENCE

DEPARTMENT OF PSYCHOLOGY

by

RITA EVELYN ALDRIDGE

FEBRUARY, 1961

ABSTRACT

This research was designed to determine the validity of a Self-Administering Short-form of the Wechsler-Bellevue Adult Intelligence Scale as a predictor of academic success. The VICS, which consisted of an assembly of the Vocabulary, Information, Comprehension and Similarities sub-tests from the Verbal Scale of the Bellevue (Form I), was the short-form test used for this investigation. It was hypothesized that the VICS was a better indicator of academic success as measured by the Registered Nurses examination scores than two other longer tests of intelligence: the American Council on Education Psychological Examination (1948 edition) and the Wechsler-Bellevue Full Scale.

The sample selected for study consisted of two consecutive classes of nurses in training from the University of Alberta Hospital. There were sixty-one students in this total group. They were contacted initially during their probation year. These students were retested on the VICS in their final year of training in order that a check of reliability as well as validity on this short-form Bellevue test could be made. The Pearson Product Moment correlations were computed to determine the validity and reliability coefficients.

Throughout this investigation, lowered correlation coefficients were obtained. Further calculations indicated that the homogeneity of the sample employed had a direct bearing on the

lowering of these correlations. In view of this, only tentative interpretations of the results were made.

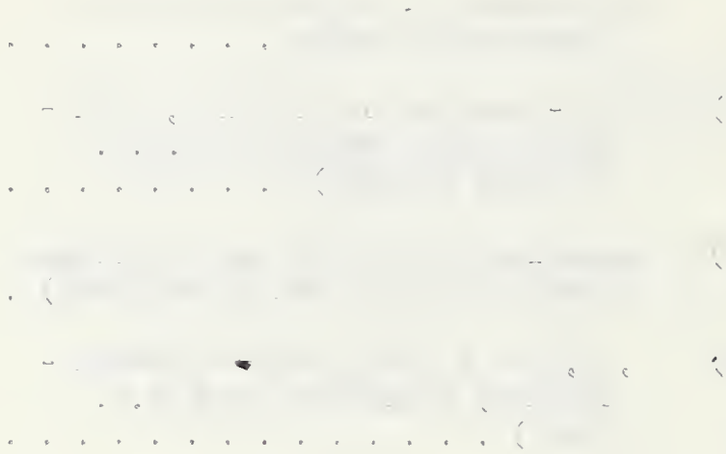
The results revealed that the VICS was not a better indicator of R.N. marks than the A.C.E. or the Full Scale. The VICS correlated highly with the other two longer tests but the coefficient obtained was lower when this short-form test was correlated with the criterion of success. The VICS was found to be a reliable test as a whole and the effect of change in method of administration did not seem to affect the reliability.

It was concluded that the use of this short-form for purposes of prediction of R.N. scores on this pre-select group of student nurses is not warranted.

ACKNOWLEDGMENTS

I wish to express my appreciation for the assistance and encouragement received during the preparation of this work. I would like to thank Dr. D. Spearman of the Department of Psychology for his advice and consultation without which this thesis would not have been completed. I am indebted to Miss L. Wilson of the Department of Psychology for her collaboration in obtaining the initial data for this investigation and for her constructive suggestions throughout this thesis work. Acknowledgment is extended to Dr. A. J. Cook of Student Counselling Services for his permission to use the necessary test data and for his assistance during the initial formulation of this investigation. Finally, I wish to express my gratitude to Miss J. Clark of the University of Alberta Hospital for her help in making the necessary arrangements for testing and for her co-operation in matters pertaining to the University Hospital Nurses Training School.

TABLE OF CONTENTS



LIST OF TABLES

LIST OF FIGURES

Chapter I

GENERAL CONSIDERATIONS

Within the past fifteen years a portion of work on the Wechsler-Bellevue Adult Intelligence Scale has been directed toward the search for shortened forms of this test. For use in the rough screening of human intelligence, high validity coefficients have been obtained on almost all combinations of Wechsler-Bellevue sub-tests. Of course, no rapid test was found to measure exactly as well as the Full Scale but some reasonable justification had been found for the use of specific short-form combinations under certain conditions.

Some general conclusions regarding these short-form investigations have been reviewed in three major publications (24, 25, 36, 39, 40). A brief note on these is as follows: (1) Short-form tests were found to be useful when a rapid estimate of intelligence was required. (2) Certain sub-test assemblies were found to be more ideally suited for use with specific subject groups. Higher validity coefficients were obtained in these cases.[1] (3) No sub-test combination offered the full diagnostic information which could be obtained from the Full Scale.

This present project is more directly concerned with the above noted second point. As will become more apparent in the detailed review

---

1. At first statistical methods employed in each of these investigations were attributed to be the cause of the discrepancy in results. However, the consistency with which this phenomenon occurred seemed to indicate that some sub-test combinations were better measures of intelligence than others.

of the various investigations of shorter forms which follows in Chapter II, consistently higher results have been found when particular sub-test assemblies were adapted for use on atypical groups. In some instances, other investigations conducted on different samples, using the same sub-tests, revealed positive but lowered results. The findings seemed to suggest that valid short-form tests of intelligence for each type of clinical population might well include a different combination of sub-tests. The search for the right combination was then usually directed toward the selection of the best sub-test combination for use with specific groups. It is notable that comparatively few such investigations have been conducted on the use of shorter-form tests on a student population. Some indication of successful rapid tests for use with student populations can be derived from the studies of college students made by Patterson (36) and Bay and Berks (3) and a much earlier investigation by Rabin (38) on student nurses. Promising results in these cases were revealed on a predominantly "Verbal" selection of sub-tests.[1] Unfortunately, the limited extent of research directed to such a group does not allow conclusions about which of these sub-tests might be most useful. It is apparent then, that a more extensive exploration into this area is required. In view of the high correlations obtained with Verbal short-form batteries on college students and nurses in training, it is suggested that applications of Verbal sub-test combinations might

---

1. Rabin selected a short-form assembly of the Comprehension, Arithmetic and Similarities sub-tests; Bay & Berks selected Comprehension, Similarities and Digit Symbols sub-tests; Patterson used Vocabulary, Comprehension and Digit Span sub-tests.

be the logical first step in reference to research in this area.

In addition, it is interesting to note at this point that Wechsler himself proposed a method of shortening the Full Scale when he suggested that the Verbal Scale sub-tests alone were probably better indicators of a rapid assessment of general intelligence for the population as a whole. He emphasized the particular usefulness of these sub-tests when he stated:

> Where the main aim is to obtain a reliable intelligence rating quickly, an abbreviated form of this test will suffice in most cases. For such purposes the simplest and safest procedure is to use the Verbal part of the examination alone. (49) p. 145

In this project, a slight departure from previous investigations will be attempted. Specifically, the administration has been changed in order to fulfil a practical need which arises in most student counselling centers. The Wechsler-Bellevue Full Scale has been a major offender in contributing to an over-balance in timing of an interview schedule.[1] Since in a great many cases only a rough estimate of intelligence level is required, a shorter form would be particularly useful. To overcome the time problem by minimizing administration time, a Self-Administering Verbal Short-form of the Wechsler-Bellevue was developed. To date there has been no report in the literature of such work. Apart from the obvious advantages in the Self-Administering Short-form there are certain other valuable features. For example,

---

1. Form I of this test is readily used as a measure of general intelligence and for the qualitative factors it reveals. However, often an hour or more is required for test administration alone (depending on the variables in the individual cases).

the Self-Administering Short-form can be used with groups and it is entirely possible that self-administration increases objectivity.[1]

Another point of interest in reference to the criterion for validity is worth considering. Recent publications (1, 32, 41) on the validity of the Wechsler-Bellevue Full Scale with college students have proposed the use of academic success as a criterion for intelligence in preference to other tests of general mental ability at this level. Their findings revealed a highest correlation with this criterion on the Verbal Scale of the Bellevue test. These results seem to further justify the usefulness of the Verbal Scale, particularly with college students.

Although the sample in this study is not specifically a college group, it is certainly a student group. In some ways the sample of student nurses in this study could be considered comparable. The requirement for entrance to nursing school is much the same as for University entrance.[2]

Problem

The aim of this research project is then, to determine the validity of a Self-Administering Verbal Short-form of the Wechsler-Bellevue Adult Intelligence Scale.

More specifically, this study will be directed toward the

---

1. Gurvitz (18) p.727. In a later paper Cohen (7) discusses the matter further and suggests that objectivity is not affected by examiner administration.

2. High School matriculation was required for entrance to either the University of Alberta or the training school at the University Hospital.

capacity of such a short-form to predict the academic success of a group of student nurses in relation to other frequently used tests of intelligence.

This may be stated in hypothetical form. A Verbal Self-Administering Short-form of the Wechsler-Bellevue composed of Vocabulary, Information, Similarities and Comprehension sub-tests, is a more accurate predictor of graduation averages of student nurses than (a) the Wechsler-Bellevue Full Scale or (b) the American Council on Education Test.

Chapter II

SURVEY OF THE LITERATURE

This chapter will include a detailed review of investigations which relate to the present project. Consequently, some of the background of short-form tests on the Wechsler-Bellevue Scale will be considered. This will be followed by a more thorough discussion of the various short-form batteries appearing in the literature. Since this project is designed to check on the validity of a short-form test for use with student groups, emphasis will be placed on a review of research conducted with short-forms upon such groups. In addition, some of the pertinent studies relating to the validity of the Wechsler-Bellevue Full Scale on this group will be mentioned. Finally, a brief discussion of the consistency of measurement of the various sub-tests will be included in the section relating to reliability.

Introduction

During World War II widespread military conscription and recruiting intensified the need for the development of a valid and rapid evaluation of intellectual function. The budgeting of working hours became important as the demands on psychological services increased. Routine tests were needed which could be administered readily for adequate initial judgements. Unfortunately, little time was left for the development of new tests. Instead, psychologists turned to abbreviated forms of the longer contemporary tests.

The Use of Short-Form Tests in Clinical Practice: As research progressed in this area, a parallel series of controversial articles was published in reference to the advantages and disadvantages of such a technique in

testing. Most of this research was in reference to short-form tests of the Wechsler-Bellevue but the resulting conclusions were applicable to all short-form measures in clinical practice. One of the major critics of short-form tests was Patterson (35, 36, 37). In his many articles on this topic he questioned the advisability of using abbreviated tests in preference to the longer established tests in clinical practice. He feared that undue emphasis might be placed on the use of these tests if the limitations of rapid measures were not fully realized by clinicians and mentioned the possible dangers which might ensue with over-confidence regarding the use of these tests. Short-form tests, he continued, were not considered to be as capable of fine discrimination of intellectual functions as the longer tests. To explain this, Patterson placed particular emphasis on the problem of enlarged errors of reliability which might occur with the construction of short tests. Further, there was decrement in the reliability of the sample of the subjects' behavior under observation. Hence, not only did he consider short-forms to be statistically questionable only as a rough measure of intelligence at best, but these abbreviated forms also placed limitations on the amount of clinical material that could be derived. Patterson concluded that the use of shorter-forms was not sound clinical procedure. Similarly, in his Fourth Edition (50), Wechsler, like Patterson, stressed that not fewer but more tests were needed in clinical practice.

On the other hand, one of the advocates of short-tests was Hunt (24). Although Hunt accepted Patterson's arguments to a degree, he felt that abbreviated testing definitely had practical value in some clinical situations. He wrote that in many cases Patterson's arguments

were inconclusive, and contended that the merit of these shortened forms for rough screening could not be denied, particularly in cases where only a rough estimate of intelligence level was needed. He agreed to the fact that shorter tests have a reduced reliability but stressed that such a difficulty was to be encountered to a lesser or greater degree in all tests, long and short. In reference to the impoverishment of diagnostic material, Hunt referred to the lack of research in this area. He implied that favourable results might be obtained but that no careful study had been completed to uphold this view. Patterson's criticism in this area, he concluded, may have been somewhat premature.

Eventually, a compromise was reached between the two schools of thought. First, short-forms were considered to be the most immediate solution to the problem of lightening the burden of routine testing. However, they should not be used in preference to longer tests if time is available. Second, sound clinical background is necessary for interpretation of these tests. Particular awareness is needed in dealing with the limitations incurred in validity and reliability. Third, if a short-form test is to be employed, it should be a part of a longer test. Thus, in the event of any abnormalities of poor performance, the testing could be continued. An additional desirable feature was mentioned in that short tests should be extracted from a longer test which is known to be capable of revealing useful clinical data.

The Wechsler-Bellevue as a Basis for Short-Form Tests: The Wechsler-Bellevue Adult Intelligence Scale is one of the longer tests most frequently used in clinical situations in the evaluation of intelligence. The physical construction of this scale is such that it lends itself

readily to shortening of administration time. Certain properties of this test suggest easy conversion into short-form. This Bellevue Scale is a point scale consisting of eleven sub-tests. The six sub-tests which Wechsler calls collectively the "Verbal Scale" are entitled Information, Comprehension, Digit Span, Arithmetic, Similarities and Vocabulary. The Vocabulary sub-test was originally included as an alternate sub-test, but the importance of its inclusion in the regular test battery has been mentioned by many authors (10, 23, 27, 37, 40). The "Performance Scale" includes Picture Arrangement, Picture Completion, Block Design, Object Assembly and Digit Symbols sub-tests. A combination of all these sub-tests constitutes the "Full Scale". Wechsler (50) converted the raw score of each of these sub-tests to a standard scale. This enabled a direct comparison of the different sub-units on an equal basis. Thus, it is not surprising that the search for an adequate short test of intelligence ended in the extraction of the sub-tests which could be combined into sub-scales to form rapid evaluations of intelligence. In recent testing the Wechsler-Bellevue Full Scale Form I has been used as one of the good whole tests of intelligence. In fact, its very rationale considers the "global" aspects in measuring intelligence over and above an addition of special abilities.[1] Although each test was to measure a special ability, in addition the individual sub-tests demonstrate the property of measuring general intelligence by

---

1. Wechsler (50) p.8.

their positive correlations with the Full Scale.[1] It is not surprising then, that the short-form investigators would reason that fragments of the Full Scale alone might measure as adequately or almost as adequately as the Full Scale. This contention was ventured by Rabin (38) in the following statement:

> It is possible that some sub-tests tend to reveal the trend of general mental level more readily than others. (p.321).

Conversely then, some sub-tests were considered to contribute less to the total intelligence score and, in turn, were less useful for inclusion in test batteries. Since most of these investigations were concerned with rapid testing for clinical, usually atypical groups, it was assumed that by selecting only those sub-tests which showed higher correlations with the Full Scale that a valid measure of intelligence could be obtained regardless of the abbreviation in form.

Short-Form Investigations of the Wechsler-Bellevue

The first short-form publication of the Wechsler-Bellevue Scale Form I was presented by Rabin in 1943.[2] Since then various individually proposed scales (3, 8, 9, 11, 15, 18, 23, 25, 27, 33, 37, 45), some general reviews (21, 22, 33, 35, 39, 47), and statistical studies (22,30) have been published. Although proportionately less research in this area has been conducted recently, some investigations

---

1. Although there is a considerable range (.409 - .727) in the correlations Wechsler obtained between sub-test and Full Scale, a significant positive correlation was obtained in all cases (Table 43). p.224.

2. Other studies have been reported on the Wechsler Adult Intelligence Scale (W.A.I.S.) and the Wechsler Intelligence Scale for Children (W.I.S.C.).

have been continued to the present date (33). In most of the short-forms, more than one author has investigated the validity. Almost all combinations of sub-tests have been used.[1] Most of the authors selected sub-tests from both the Verbal and Performance sub-scales. Some chose tests from the Verbal section only (11, 27, 38). One (25) inserted verbal items from another test in his battery.[2] Test administration time varied from five minutes (18) to thirty minutes (15, 27, 37). A listing of these early major investigations on abbreviated forms appears in detail in the reviews of Rabin (39) and Patterson (35).

In addition, more comprehensive surveys of essentially a statistical nature were conducted by three authors on more heterogeneous populations. McNemar (29) studied the ten best combinations of one to six sub-tests. He devised an economical statistical technique which made it possible to study many combinations on Wechsler's standardization groups.[3] In a later study, Herring (21) investigated all the combinations of sub-tests to the time of his study, using a large group of normal and abnormal white male veterans. Hilden (22) used a similar procedure to that of McNemar and reported on an extensive analysis of all combinations of sub-tests on a large population of Psycho-Neurotics of similar age

---

1. The Vocabulary sub-test, considered as an alternate in the earlier publications, was not included in many of the studies.

2. The Thorndike Verbal Scale was included with the Bellevue Comprehension and Similarities sub-tests. The author converted this new scale to the similar standard score of Wechsler.

3. McNemar employed a modified correlation technique to determine the extent to which different combinations correlated with the Full Scale. Wechsler's data was adopted for these inter-correlations.

range to Wechsler's normative data.[1] In this same investigation, Hilden reviewed and summarized in great detail the methods adopted in the many individual investigations on short-form Bellevue tests. Under many sub-headings he included a thorough discussion on the various methods of selection of sub-tests, criteria used for validity, populations selected for study, and the various age and intelligence levels of these samples. Emphasis was placed on the lack of co-ordination of research among these various investigations as well as inconsistencies in reporting of pertinent data.

Results on Short-Form Investigations: The results from the various short-forms looked promising. In most cases validity coefficients with the Full Scale ranged from .80 to .90.[2] Remarkably enough, all investigated test combinations were considered to be useful for rough screening purposes. However, consistently higher results were obtained from the individual investigations where certain sub-test combinations were selected for use with atypical groups only. It seemed that in these cases the chosen sub-test groupings were more sensitive for intellectual discrimination of general intelligence on these select samples. This was particularly noticeable when the same test was used on different samples by others (5, 9, 21, 24, 35). In this case the correlations were slightly lower,

---

1. Inter-correlations were computed and a correction formula was applied to the standard deviation for estimating results. By this technique the sample was made to be more representative of the Wechsler normative data.

2. Adequate comparison as to the coefficients alone is difficult. In these investigations the various "r's" were obtained on different criteria. In addition, the inconsistencies in using I.Q., weighted score or single sub-test scores in the analysis, should be noted.

which suggested that the test might not be as applicable for use with another group. The empirical studies conducted on more heterogeneous populations revealed similarly lowered but significant validity coefficients. Many enthusiastic claims were made from these results. It was speculated that Wechsler's sub-tests were such that certain groupings measured "special abilities" which together constituted a measure as comparable as the Full Scale in measuring general intelligence factors. Thus, Object Assembly and Picture Arrangement were found to be the least useful for use in shorter form on any population. On the other hand the importance of the Vocabulary sub-test for most groups was mentioned by Patterson (37), Hunt (25), Kriegman (27), and Hilden (22). Other findings in reference to short-forms questioned the use of Arithmetic and Digit Span sub-tests for normal subjects on the basis that they induced anxiety for the examinee and consequently inhibited the performance, (3, 37). Generally, the Verbal sub-tests were found to correlate best with the Full Scale on groups of above average intelligence whereas the inclusion of more Performance sub-tests seemed desirable for use with borderline cases. As a result, emphasis began to be placed much more on the best selection of the best sub-test combination rather than good research.

As could be expected, criticisms of these conclusions soon appeared in an article by McNemar (30) and a more recent review by Hilden (22). McNemar cast some doubt on the high validities of the results obtained on the basis that too heterogeneous groups were used in their analysis when compared to his standardized sample. He suggested that the lowered correlations obtained on his group was due

only to statistical difference of the structures of the samples selected for study rather than the measuring qualities of the sub-tests alone.[1] Hilden agreed with McNemar and maintained that these discrepancies in correlations could arise from variations in the character of the group. He particularly emphasized the differences in mean level of the population samples used as well as the great variance in standard deviations.[2] Moreover, he added other factors which could account for the "spuriously" high coefficients obtained. He stated that the raised validity coefficients in short-form studies as a whole depended in most cases on the short-form properties themselves. He mentioned the degree of intercorrelation among the tests and cited the danger of a short-form being correlated against the Full Scale of which it is a part. He added further, that the effect of an increase in the number of sub-tests included in these assemblies would proportionately increase the final correlation.

Gradually, as more statistical research on short-forms was attempted, the entire area fell into disfavor as controversy over methodology came to the fore. Particular criticism was directed to the manner in which individual reports were investigated without any co-ordination of research. Specifically the lack of completeness in what was reported was criticized.[3] Thus, most of the claims of short-form investigations were carefully reconsidered and the importance of short-form tests began to diminish. In fact, the amount of time devoted to this topic was

---

1. McNemar (30) p.80

2. He considered the dispersion of I.Q. with Wechsler's theoretical S.D. of 14.83. Although very few investigators reported the S.D.'s, the range with the Full Scale varies from 7.66 to 20.66.

3. Hilden presents these arguments in his article (22) p.329.

considered to be grossly disproportionate in relation to the importance of this area in clinical testing (35). However, the clinical value of short-form tests for rapid assessment of intelligence was accepted for rough screening purposes.

Short-Form Studies of the Wechsler-Bellevue on Groups of Above Average Intelligence

As has been mentioned before, most of the short-forms were designed to be used in the busy clinical situation, where the time element was an important consideration. Strangely enough, few such investigations have been reported on student groups. Three investigations of this type have been mentioned. Rabin (38) selected a group of 92 student nurses with an I.Q. range of 80 - 130. The mean and standard deviation were not reported. He obtained a coefficient of .80 with the Full Scale I.Q. on a three Verbal sub-test battery consisting of Comprehension, Arithmetic and Similarities sub-tests. The two other references are found in the literature dealing with short-form tests on college students. These studies were not directly concerned with these groups but included them as part of a larger study.[1] In an attempt to include the Vocabulary sub-test in a short-form combination, Patterson (37) found that a Vocabulary, Comprehension and Digit Span assembly yielded a high validity coefficient. This was particularly noticeable in the higher I.Q. levels of his male veteran sub-group. Bay & Berks (3) studied the validity of an assembly of Comprehension, Digit Symbols and Similarities on a group of 48 male college students.

---

1. College students were used in sub-groups or cross validation studies.

The correlation with the Full Scale was established at .76. In all these studies the correlations obtained with the shorter-forms were significant. The predominance of Verbal Scale sub-tests is also general. Only the study by Rabin, however, can be considered comparable to the present investigation, since a group of student nurses was employed for study in both cases.

An allied series of articles concerned with the validity of the full Wechsler-Bellevue Scale on college students has recently presented some new views relating to Verbal short-form tests. Although this research was concerned primarily with a higher intellectual group, a review of their findings is of interest. An attempt was made to determine the effectiveness of the Wechsler-Bellevue as a measuring instrument of intelligence on college students. Anderson (1) and Merril (32) adopted a new criterion for validity which was considered to be more applicable for academic purposes. They assumed that when "intelligence" was to be considered at the college level, emphasis should be placed on prediction of success in preference to the general intelligence of the population as a whole. When they compared the Wechsler-Bellevue Scale and other tests of general mental ability with this criterion, the best correlation with Grade Point averages was found in the Verbal Scale alone. These findings suggested the superiority of this portion only of the Bellevue Scale for use with this group. Merril concluded:

> This scale provides the counsellor with the opportunity to observe the quality of the students behavior ..... One can speculate that academic predictions based on the Verbal Scale are best because it reflects like the Grade Point averages both personality and ability factors. (32) (p.65).

It is possible then, that in other situations where academic performance is stressed the use of the Verbal Scale may be the best clinical indicator of success in these fields. Inasmuch as this research is concerned with Verbal sub-tests only, some support can be gained from the above work for the present study.

Reliability

Few reports in the literature are found concerning the topic of reliability studies of the Bellevue Scale. Only one investigation into the reliability of a short-form appears to date (27).[1] Perhaps the reliability investigation by Gilhooly (17) can be considered to be more directly related to the present project on this badly neglected area. The Odd-Even split half reliability coefficient was calculated in this case on the Information, Comprehension, Similarities and Vocabulary sub-tests. Highest degree of variance was found with the Comprehension sub-test and the Vocabulary was found to be the most stable. Further support of the reliability of these four sub-tests can be derived from the Full Scale reliability studies: test-retest (2, 12, 20, 41, 49), split half (17, 48), and alternate form methods (16, 39, 50). Interestingly enough, results obtained on these various Full Scale studies seemed to be in general agreement with the findings of Gilhooly. Considering the four sub-tests of interest in this study, Vocabulary and Information were found to be consistently more reliable than the Similarities and Comprehension sub-tests.

---

1. Kriegman & Hansen adopted an Odd - Even split half technique on their short-form of Information, Vocabulary, Similarities and Block Design.

Summary

In summary, a great deal of research on short-form tests of the Wechsler-Bellevue has been conducted. Results have been promising for rough screening purposes. Any definite conclusions about particular sub-test combinations are difficult to form since there has been little consistency in technique for dealing with sample and data. More careful co-ordinated research in this area is to be desired.

Although no rapid test measured as adequately as the Full Scale, certain combinations were found to show good correlations on specific groups, despite their abbreviated form. Recent empirical investigations have questioned these "spuriously" high validity coefficients. However, the fact that fairly good results were obtained consistently in these specific individual investigations, suggests that these test batteries have some clinical use.

Little has been done in investigations concerned with short-form studies on student groups. Only one study, on student nurses, directly relates to the present project. A Verbal sub-test battery was employed in this instance. Recent validity studies regarding the use of the Bellevue Scale for college students have found the Verbal Scale to be most useful for academic purposes. In addition, the consistency of measurement of this scale has proven to be high from the various individual investigations, although considerable variance has been noted among individual sub-tests.

All of these findings agree on the Verbal Scale as being the most reliable and promising portion of the Bellevue Scale for measuring intelligence. More recent reports suggest that it may be even more

applicable for use with student populations. Thus, some support is obtained for a preliminary investigation relating to the assessment of intelligence of a student group by means of a self-administering short-form employing sub-tests from the Verbal section of the Bellevue.

Chapter III

PROCEDURE

This chapter will present a description of the Self-Administering Short-Form and its administration and scoring. The various procedures adopted for purposes of study and an analysis of the sample under investigation will be described in some detail. Finally, there will be some considerations of the statistical methods to be employed in working with the data.

Description of the Test

The Self-Administering Short-Form test developed for this study consists of a four test assembly of Verbal sub-tests derived from the Wechsler-Bellevue Full Scale. The individual sub-tests employed are Information, Comprehension, Similarities and Vocabulary. For purposes of brevity the method of notation used by Rabin and Guertin (39) will be applied in referring to these sub-tests.[1] Thus, I, C, S and V will be used for Information, Comprehension, Similarities and Vocabulary respectively. In order to simplify discussion the entire test will be abbreviated to the VICS.[2]

These sub-tests are presented in the sequence order selected by Wechsler. Thus, this test is divided into four sections, each representing one of these sub-tests. They are in mimeographed form,

---

1. Rabin & Guertin chose double letters to represent Performance Scale sub-tests and single letters for Verbal Scale sub-tests.

2. This test was originally devised by Dr. D. Spearman, Professor of Psychology in the Department of Psychology, University of Alberta.

stapled together to form a booklet. Since this form is self-administering, directions are written upon the sheets rather than presented orally.[1] In the original test ample space is provided at the end of each question to enable the examinee to place his answer in writing. There is no time limit on this test. A copy of this self-administering form appears in Appendix A.

Selection of Sub-Tests: The rationale for the selection of these four sub-tests is dependent upon three main factors. First, they are all Verbal sub-tests. This is known to be useful for this specific group.[2] Second, they are the only ones which lend themselves readily for self-administration. Third, they are for the most part among the best tests in the Wechsler group.[3] In addition, limiting the short-form would seem

---

1. Some changes in the test items for Canadian subjects have been made.

   (a) Changes in the information test were made in the following items:

   (a) Who is the President of the U.S.?
   1. Who was the President before him?
   9. How tall is the average American woman?
   15. What is the population of the U.S.?
   16. When is Washington's birthday?

   Replacements were:

   (a) Who is the Prime Minister of Canada?
   1. Who was the Prime Minister before him?
   9. How tall is the average Canadian woman?
   15. What is the population of Canada?
   16. When is Queen Victoria's birthday?

2. The matter of Verbal sub-test selection for student groups has already been discussed in Chapter I.

3. These tests are among those which correlate the highest with the Full Scale.

to follow the thinking of Hilden (22).[1] A compilation of the especially valuable aspects of intelligence which each of these sub-tests measures, appears in a more theoretical article by Watson (47).[2]

Administration: The administration involves giving the test booklet to the subject. This may be done for individuals or in group form. The subjects are provided with pencils. Written instructions printed on the first page are as follows:

> Please answer the questions on the following pages. Enter your responses in the blank spaces provided. Write as clearly as possible. Work as quickly as you can.

It is sometimes helpful to read these instructions aloud to the subjects. This has been the approach used in this study. Further instructions relating to the specific sections of the test are also printed and precede each series of sub-test questions. As will be noted, some modification in Wechsler's standardized instructions have been made to adhere to a pencil and paper examination. However, an almost negligible change in this presentation will be noted. For Section I and Section II, which include the Information and Comprehension items, the instructions are simply:

> Answer these questions.

The instructions for the Similarities sub-test are more explanatory. They are as follows:

---

1. Hilden considers that more than four sub-tests in a test battery lose their effectiveness and contribute proportionately less to the total score than would warrant their inclusion.

2. Watson discusses the views of Wechsler, Rabin and Rapaport on this subject.

> In the list below you will find three things which are the same or alike in certain ways. You are requested to write down in what way they are alike. For example: 'In what way are an orange and a banana the same?'
> For this you might say: "You can eat them both' or 'They both have skins' or 'They are both fruits'. They are alike in three ways. Write down ONE way in which they are alike.

Finally, the written instructions for the Vocabulary sub-test are:

> Give the meaning of each of the following words.

There is no strict time limit on this test. (The subjects are asked to work as quickly as they can). Some note of the test time is recorded.

Scoring: Scoring is done in a manner similar to that appearing in Wechsler's manual (49).[1] The weighted scores on these four sub-tests are totalled and pro-rated.[2] These tests may be converted to I.Q. scores by using the Wechsler Verbal I.Q. table (49).[3]

Method

The American Council on Education Test (1948 Edition) and the

---

1. Modifications in scoring for Canadian subjects have been made.

2. That is 1/4 of the additive value of the weighted scores on Information, Comprehension, Similarities and Vocabulary, multiplied by 5. The formula being

$$5\left[\frac{I + C + S + V}{4}\right]$$

3. Previous pro-rating techniques by Rabin, Geil, Springer and Hunt were based on Wechsler's contention that each sub-test contributed equally to 1/10 of the Total Score. However, Kriegman and Cummings employed a multiple regression technique and found that each sub-test was not 2.5 of the whole. As a result, they established a new coefficient of 2.7 for their purposes. Realizing this difficulty, the remainder of the investigators preferred to use weighted scores only.

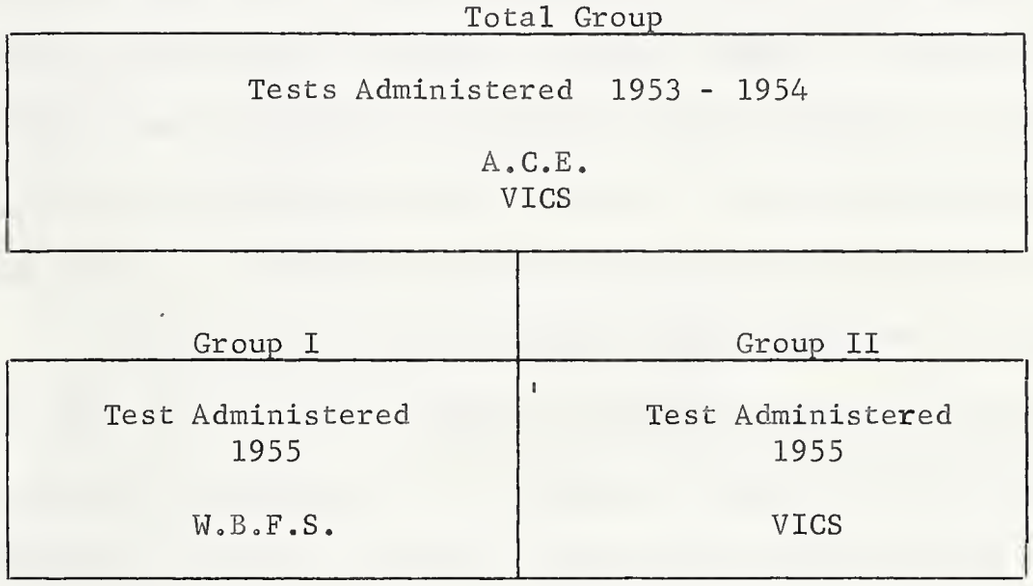
VICS were administered in group form to two consecutive classes of student nurses who were registered in the Diploma Course of Nursing at the University of Alberta Hospital.[1] These tests were given upon admittance along with the regular battery of entrance examinations by a staff member of Student Counselling Services, University of Alberta.[2] It is of interest to note that the average test time for these groups on the VICS was 38 minutes. These results were scored and remained on file for two years.[3] At this second time the entire group was again contacted. The research project was explained to them and all volunteered to be tested. This total group was then divided into two to facilitate a study of reliability and validity of the test. Group I was contacted and arrangements were made for Wechsler-Bellevue Full Scale testing.[4] In addition, listings of these appointments were placed on the Bulletin Board in the Nurses' Residence. Testing of Group I was scheduled within a three-month period from October, 1955

---

1. The University of Alberta also offers a degree course in nursing in conjunction with the University of Alberta Hospital.
2. Miss Lolita Wilson, Assistant Director of Student Counselling Services and Associate Professor in the Department of Psychology, University of Alberta.
3. The A.C.E. was machine scored in all cases.
4. These appointments were made with the co-operation of the Nurses Teaching Staff and the Class President. A copy of the form letter employed is included in Appendix B.

to December, 1955.[1] The average test time for Full Scale administration was approximately 60 minutes. At a later date, the remaining students in Group II were re-tested on the VICS in group form. The design for testing was as follows:

Figure 1



Design of the Experiment

Description of the Sample

Originally, the entire class of 69 student nurses was included for study. However, during the course of this long range

---

1. Testing was done by the author and another graduate student, Mr. Lorne Kendall, Graduate Student in Psychology, University of Alberta. With the exception of changes for the benefit of Canadian subjects, Wechsler's instructions for administration and scoring were adhered to in dealing with Full Scale testing. Further checks on scoring were conducted on the Full Scale items and the VICS items with the assistance of Dr. Spearman, a member of the Department of Psychology. Questionable answers on the Verbal Scale were discussed for scoring. Comprehension and Vocabulary items were particularly scrutinized. The remainder of the protocals were spot checked one in three.

research period, seven of these students withdrew from training.[1] The remaining 62 students constituted the subjects for study for this investigation. All these subjects were white female with the exception of one student.[2] Socio-economic status was random from the Prairie Provinces with an almost equal proportion of rural and urban subjects. Entrance requirement in this nursing course was a Senior Matriculation average in High School of 50%. The small number of out-of-province candidates were admitted on their High School averages.[3] Figure 2 reveals the distribution of this group on the basis of their High School marks. The shaded portions represent out-of-province students.

An attempt was made to match these groups on the basis of their High School marks. Unfortunately, substitution occurred without the examiner's knowledge.[4] In addition, another difficulty was encountered in the final stages of testing when one member of a matched

---

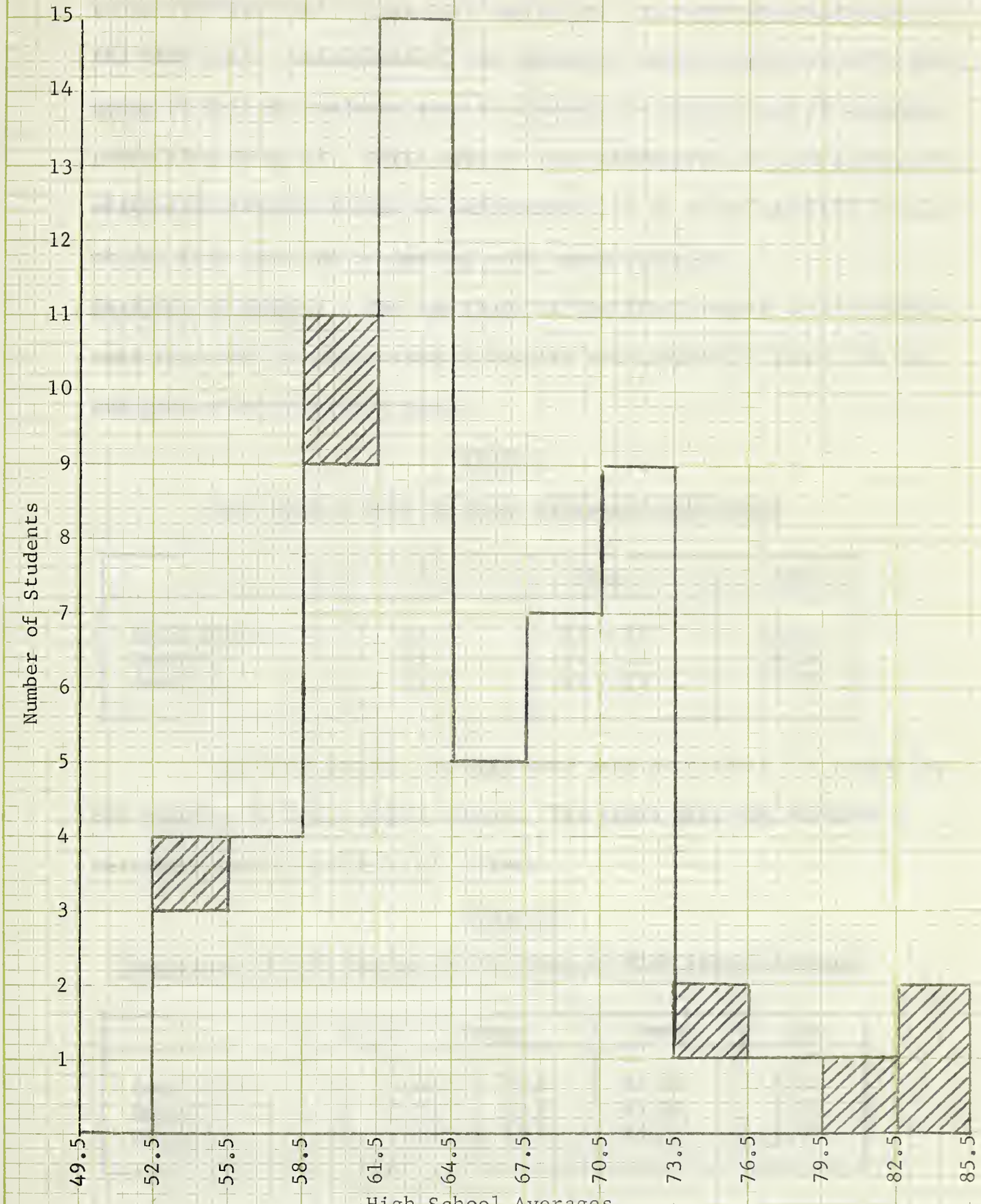
1. In the first few months after registration, one student changed to an Arts and Science pattern, another failed her probation period and five withdrew to be married.

2. This student was of Japanese descent but had resided in Alberta since childhood.

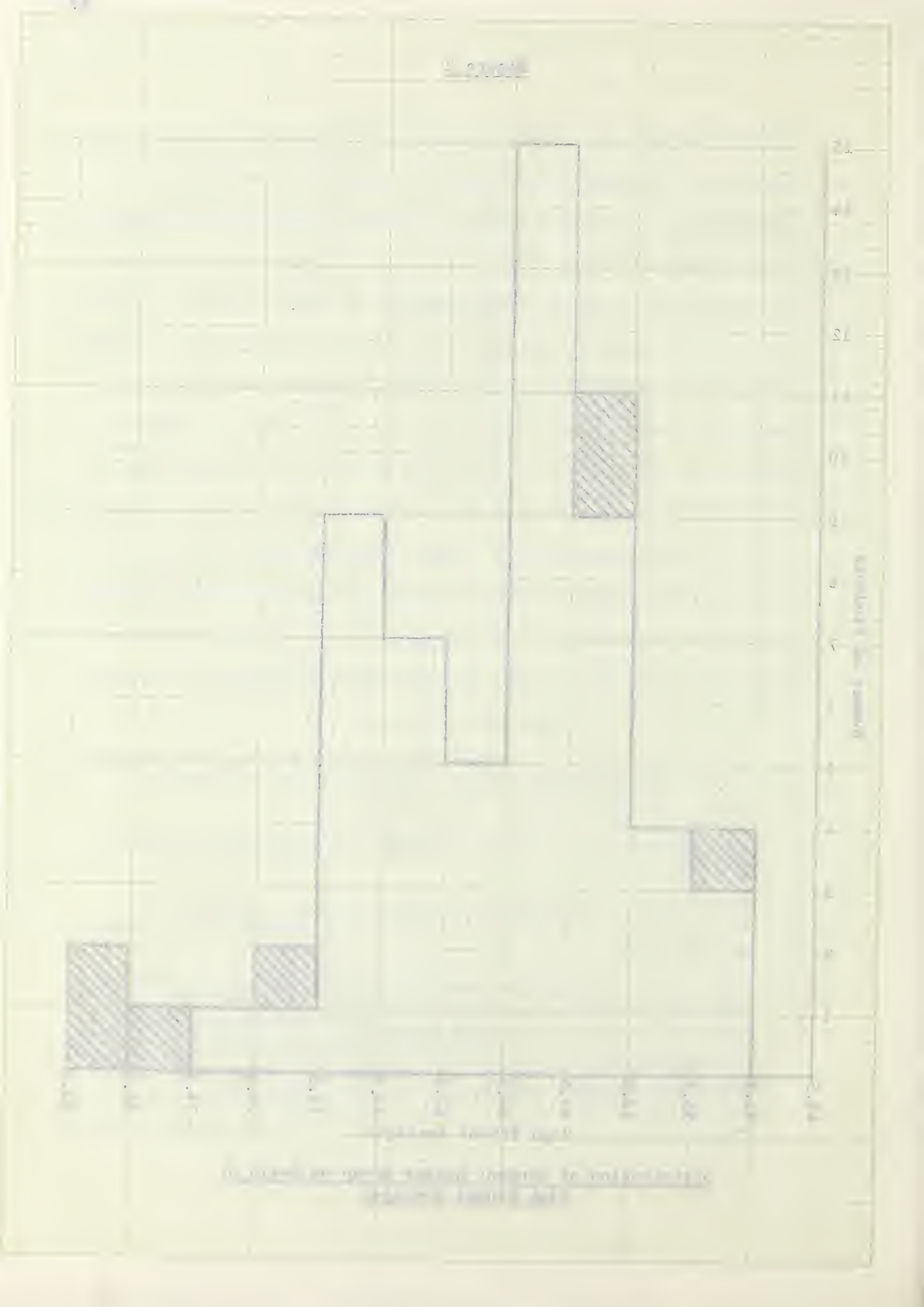
3. The question of admittance of out-of-province candidates is determined by the Registrar's Office, where equivalent standings are accepted.

4. Unavoidable substitutions occurred because in many cases shift work conflicted with the time of test administration. Thus, afternoon and night shift students were predominant in Group I. At the time of group testing, no such difficulty arose since Group II students were again attending classes. An analysis of these various shifts appears in Appendix B (ii).

Figure 2



Distribution of Student Nurses Group on Basis of High School Averages



group withdrew from nurses training before the final administration of the VICS test. Consequently, the remaining sample consisted of a total group of only 61 subjects with 31 students in Group I and 30 students comprising Group II. While some of this difficulty in maintaining the originally matched groups is unfortunate, it is to be expected in view of the fact that one is dealing with human subjects.

Analysis of Sample: The age range of the final sample of 61 students used for study in this research project was comparable among the two sub-groups and the total group.

Table I

Age Range & Mean of Total Group and Sub-Groups

	N	Range	Mean
Total Group	61	18 - 22	18.61
Group I	31	18 - 22	18.87
Group II	30	18 - 19	18.33

The High School averages were used as a basis for comparing the sampling of these three groups. The range mean and standard deviation were calculated as follows:

Table II

Comparison of the Sample on the Basis of High School Averages

	N	Range	Mean	S.D.
Total Group	61	53.6 - 85.2	65.63	6.94
Group I	31	55.7 - 83.2	65.65	6.40
Group II	30	53.6 - 85.2	65.61	7.47

The "t" critical ratio for the significance of difference between the means of Group I and Group II was computed.[1] The resulting "t" was .552 which showed no significant difference between these groups. The assumption can therefore be made that these two groups represent the same population.

A frequency polygon was also plotted to enable an inspection of the type of distribution obtained in these groups. A slight skewness and restriction of range could be noted throughout (see Appendix B (iii)). However, by inspection the construction of the polygons in the three groups appear to be very similar.[2]

Procedure

To determine the validity of the Self-Administering Verbal Short-Form of the Wechsler-Bellevue Intelligence Scale on a student nurses population, the criterion of graduate marks will be used. In

---

1. The formula employed for this calculation appears in Lindquist (28) p.70.

$$t = \frac{M_a - M_b}{\sqrt{\left(\frac{\Sigma d^2_a + \Sigma d^2_b}{(N_a + N_b) - 2}\right)\left(\frac{N_a + N_b}{N_a N_b}\right)}}$$

2. The checks for normality and skewness were not employed in this instance since the advisability of using such a procedure on so small a group is questionable. (29) p.10.

addition, two other tests will be correlated for interest: the American Council on Education Test (A.C.E.) 1948 total I.Q. scores and the Wechsler-Bellevue Full Scale weighted score.[1] The Registered Nurses' examination score means[2] will be employed as the indicator of success. Along with determining the applicability of this Self-Administering form for use in clinical practice, some estimation of the consistency of measurement of this scale should be known. Thus, the reliability of this test will be computed as well as the validity. In this project the test-retest method is proposed. A two-year time interval has been adopted to ensure against any "memory carry-over". These tests are known to be fairly reliable, as has been discussed in the preceding chapter (2, 12, 17, 20, 48).

Statistical Method

The statistical procedure selected for working with the data

---

1. For purposes of discussion both the A.C.E. and the Bellevue Full Scale will be referred to as tests of intelligence. Specifically, the following explanation of what these tests purport to measure is given by the authors:

> "The purpose of the American Council on Education Psychological Examination is to appraise what has been called scholastic aptitude or general intelligence.....and to distinguish a student's mental abilities from his high school preparation and industry." (46) (p.2)

   For Wechsler's definition of intelligence see (50) p.4 et seq. Also a brief discussion of the rationale of his test is included in Chapter II of this study (p.9).

2. The Registered Nurses (R.N.) marks are the scores on the professional examinations for all nurses in the United States and Canada as presented by the National League of Nursing. (See Appendix C).

was the Pearson Product Moment Correlation.[1] This technique was used throughout for determining the relationship of validity as well as the reliability of this Self-Administering form of the Bellevue Scale. The various inter-correlations among these tests were calculated. These included the "t" test for determining the significance of the calculated "r's".[2] Significance was determined at the .05 level. In addition, some assessment of the accuracy of prediction of the R.N. marks from the validity coefficients of these three tests was made. Thus, the Index of Forecasting Efficiency was calculated in order to determine the improvement in prediction of the true criterion variables on the basis of the knowledge of these test scores over the prediction without

---

1. The formula used for arriving at the correlation is taken from McNemar (29) p.118. The machine scoring method has been employed.

$$r_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{\left[N\Sigma X^2 - (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

2. This formula was obtained from Garrett (14) p.298. The "t" ratio is:

$$t = r\sqrt{\frac{N - 2}{1 - r^2}}$$

Table 49 in Garrett, p. 464, was employed for determining the significance of the resulting "t".

knowledge of these scores.[1]

Discussion

A question had been raised during the course of research as to the advisability of using a non-parametric analysis for determining the results in this particular study. However, when the advantages and disadvantages of using this technique were carefully considered, the parametric Pearson "r" method was chosen to be the most desirable. In spite of the restrictions imposed,[2] this technique was selected because it is considered to be the most sensitive method of determining relative standings when comparing a set of individual cases on one correlated sample with another (43). Also, the fact that the majority of investigations on short-forms had adopted this same technique is another reason for the selection. Comparisons of their results with those obtained in the present project would be more meaningful since the same statistical procedure has been adopted.

---

1. The following formula is a derivative of the correlation coefficients:

$$E = 100\ (1 - \sqrt{1 - r^2})$$

This formula appears along with a discussion concerning its interpretation and use in Guilford, J.P., Fundamental Statistics in Psychology and Education, 3rd Ed., McGraw Hill Book Co. Inc., Toronto, 1956 (pp. 375-379).

2. McNemar (29) and Siegel (43) both suggest that this is a better technique, providing that the basic assumptions of normality and linearity are accepted. For purposes of this study, these have been assumed. It is felt that more valuable information would be lost by using a non-parametric "rho" rank-order correlation, than would warrant its use to overcome errors of sampling.

Chapter IV

RESULTS

A presentation and a discussion of the results will be given in this chapter.

Section I will be directly concerned with the major hypothesis as it includes an analysis of the validity of the VICS on a group of student nurses. Inter-correlations were computed between the various criteria and the Self-Administering Verbal Short-Form test. Initially, a comparison of the VICS with the criterion of success was made to determine the extent to which this shorter form predicts success in relation to the A.C.E. Both of these tests were correlated with the Registered Nurses marks. Correlations between these two tests were also calculated. The total group of 61 nurses was employed in this analysis. The scores obtained on the VICS and the A.C.E. in the 1953-54 administration were employed.

A similar procedure was adopted with the Wechsler-Bellevue Full Scale, VICS and Registered Nurses marks on Group I of the sample. The correlation between this short form and the Wechsler-Bellevue Full Scale was calculated for the purpose of comparing these present findings with the results of other short-form investigations reported in the literature.

Section II will be concerned with the test-retest reliability of this Self-Administering Verbal Short-Form test. The procedure for this analysis was to correlate the pro-rated weighted score of the VICS, administered 1953-54,and the pro-rated weighted score of the VICS,

administered 1955, on Group II of the sample. The sub-tests as well were compared individually to determine their relative stability of measurement.

In addition, and closely allied with Section II, will be an analysis of the effect of transforming these four Verbal sub-tests into a self-administering form. This will appear in Section III. The data obtained from Group I of the sample was employed to estimate whether there is any variation in consistency of measurement among the sub-tests which can be self-administered and the sub-tests which are administered by the examiner. Comparisons of the pro-rated VICS weighted score and the individual weighted sub-test scores on the Self-Administering VICS (1953-54) were made with the VICS combination and the V, I, C and S sub-tests found in the Wechsler-Bellevue Full Scale administration.[1] The relationship of Section II and Section III should exhibit some interesting conclusions. Not only is it expected that an estimate of the reliability of the individual sub-tests under different administrative conditions be revealed, but the extent to which the examiner might influence the subjects' performance by oral presentation will also be considered.

Finally, the percentage in the reduction of errors of prediction of the R.N. marks will be calculated to indicate the Index of Forecasting Efficiency of the VICS, A.C.E. and Wechsler-Bellevue Full Scale on this group of student nurses.

---

1. The same time interval of two years was employed as in Section I.

Section I

(a) Comparison of the VICS, A.C.E. and Registered Nurses Marks on the Total Group (N=61)

The correlations obtained in this section were positive but low. Data on the relationship among these tests has been considered mainly in terms of the "t" test of significance at the .05 level. The validity coefficient obtained when the VICS was correlated with the Registered Nurses marks was .20. This was not found to be significant and the probability of a zero relationship occurring in this instance was estimated as approximately 16 chances in 100.[1] The .05 level was surpassed in the correlation obtained when the A.C.E. was compared with the criterion of success. The validity coefficient revealed in this case was .38. A slightly higher correlation of .42, significant at the .001 level, was obtained between the VICS and the A.C.E. Complete data of the results in this section appear in Table III. The means and standard deviations of the tests used in this analysis, along with the Wechsler-Bellevue Full Scale mean and standard deviation on the smaller Group I, appear in Table IV.

Table III

Inter-Correlations of VICS, A.C.E. & R.N. Marks
Correlations Obtained on N=61

	r	t	p
VICS vs.R.N.	.20	1.61	-
A.C.E. vs R.N.	.38	3.17	.01
VICS vs A.C.E.	.42	3.62	.001

---

1. Garrett (14) Table 49, p.466.

Table IV

Comparison of Means and Standard Deviations

	VICS	A.C.E.	R.N.	W.B.F.S.
Mean	59.12	109.64	589.53	119.58*
S.D.	6.18	16.31	60.34	6.99*

* Group I (N=31)

(b) Comparison of the VICS, Wechsler-Bellevue Full Scale, A.C.E. and Registered Nurses Marks on Group I (N=31)

The correlation coefficients obtained from the relationship of these five tests were again low. These coefficients appear in Table V. The Full Scale revealed a correlation of .41 with the criterion of success. This is slightly higher than the correlation obtained with the A.C.E. and criterion of success in the previous section. Significance at the .05 level was established.[1] The highest coefficient was found between the VICS and the Wechsler-Bellevue Full Scale. In this instance the .42 was also significant on this smaller group.

Table V

Validity Inter-Correlations on Group I

	r	t	p
W.B.F.S. vs R.N.	.41	2.39	.05
VICS vs W.B.F.S.	.42	2.51	.05
A.C.E. vs W.B.F.S.	.41	2.41	.05

---

1. Calculations on this smaller group were repeated in the case of the VICS vs R.N. scores, and A.C.E. vs R.N. scores. These correlations were slightly lower but did not reveal a change in significance level, see Appendix D (i).

From the coefficients obtained thus far, it appears that the VICS does not predict success, as measured by the R.N. scores, better than the A.C.E. or the Wechsler-Bellevue Full Scale. No significant relationship was found between the VICS scores and the R.N. scores on a group of student nurses. The A.C.E. and the Bellevue Full Scale reveal significantly better validity coefficients than this self-administering short-form. The Full Scale presents a higher correlation with the criterion of success than the other two tests. A better degree of agreement is revealed when all three tests are compared with each other. The VICS correlates equally as well with the A.C.E. as with the Full Scale.[1] A more significant correlation was obtained between the VICS and the A.C.E. A slightly lower validity coefficient was obtained between the A.C.E. and the Bellevue Scale.

For interest, additional checks were conducted to ascertain the possibility of other factors which might have influenced this reversal of results according to the hypothesis. As can be noted, the VICS is a highly "Verbal" scale, whereas the A.C.E. and Bellevue Full Scale included Performance and Quantitative sections in addition to verbal type items. The question as to these other "performance type" factors contributing to higher correlations with R.N. scores was considered. The correlation coefficients were calculated between the Verbal and Performance weighted scores of the Bellevue and R.N. marks.[2]

---

1. For interest and for comparison of other short-form studies using I.Q. scores, the I.Q. VICS scores and the W.B.F.S. I.Q. scores were also calculated, an "r" of .31 was obtained.

2. See Appendix D (ii).

Both coefficients were low and were not found to be significant. A higher correlation was revealed in the comparison between the Verbal Scale scores and the criterion of success. It appears from this that there is an indication that the performance items in this longer test do not seem to be a major factor in prediction of success on this group.

Further, each sub-test of the VICS was correlated separately with the criterion to determine the degree of relationship each sub-test has with R.N. marks. The results of this analysis were again low and no significance of the coefficient could be derived at the .05 level.[1] Vocabulary revealed the highest coefficient of .33 with the criterion. In view of these coefficients, it is difficult to determine if any of the four sub-tests contribute to a raising or lowering of the correlations with the criterion. It seems that all these sub-tests are affected by a suppressor variable which is operating to lower the results obtained.[2]

---

1. These results appear in Appendix D (iii).

2. The Verbal C.A.S. combination of Rabin was extracted from the Bellevue Full Scale administration in this study on Group I. The obtained correlation of the C.A.S. with R.N. marks was .298. This is slightly higher than the "r" obtained from the VICS assembly; however, in both cases of these short-form tests they did not measure as well as the longer tests with the criterion.

Section II

Test-Retest Reliability of the VICS Self-Administering Short-Form on Group II (N=30)

The test-retest reliability of this Verbal Self-Administering Short-Form was found to be .57. This correlation reveals the test to be a stable measure at the .05 level. More variation in consistency of measurement was found between the reliability coefficients of the individual component sub-tests. Correlations between the Information, Comprehension, Similarities and Vocabulary sub-tests over a two-year interim range from .25 to .85. No significant differences from a zero correlation could be derived from the reliability coefficients obtained with the Similarities and Comprehension sub-tests. A greater degree of fluctuation was noted on the Similarities sub-test. The coefficients obtained on Similarities and Comprehension were .25 and .31 respectively. The correlation of .85 obtained on test-retest of the Vocabulary exceeded that of the total score. These coefficients along with the data on means and standard deviations appear in detail in the following Table VI.

Table VI

Test-Retest Reliabilities

	VICS 1953-54		VICS 1955		Correlations Obtained		
	Mean	S.D.	Mean	S.D.	r	t	p
VICS	58.70	6.22	59.16	6.18	.57	3.65	.01
I	11.36	1.76	10.86	2.45	.44	2.62	.05
C	12.26	1.57	12.73	2.08	.31	1.72	-
S	12.10	2.39	12.26	1.82	.25	1.35	-
V	11.20	3.05	12.23	2.82	.85	8.51	.001

The inclusion of the Vocabulary sub-test seems to be a determining factor in maintaining the stability of measurement of the test as a whole.

Section III

Effect of Examiner Influence on the Reliability of the Sub-Tests of the VICS Short-Form on Group II (N=31)

The reliability coefficient obtained between the VICS self-administered form and the VICS examiner administered form was found to be .35. This correlation was found to be significant at the .05 level. Correlations obtained on the Information, Vocabulary and Comprehension sub-tests were found to be significant. Information was the most reliable measure with an "r" of .73. Comprehension was barely significant with a correlation on test-retest of .47. This data along with the means and standard deviations appears in Table VII.

Table VII

Test-Retest of the V, I, C, S Examiner Administered Test and the V, I, C, S Self-Administered Test

	VICS		WBFS		Correlations Obtained		
	Mean	S.D.	Mean	S.D.	r	t	p
VICS	59.48	4.96	60.89	5.71	.35	2.52	.05
I	11.70	1.37	11.64	1.52	.73	5.77	.001
C	12.03	2.27	12.77	1.54	.47	2.84	.01
S	12.74	2.53	13.61	2.01	.13	.76	-
V	10.73	1.63	11.51	1.57	.68	4.97	.001

It appears from comparing the results on Section II and Section III that, when the method of test administration is not constant (as in Section III), a lower correlation on test-retest is revealed. The reliability of the VICS as a whole is affected somewhat where administration conditions are changed. It can be noted that there is a greater discrepancy between the means on retest in Section III than Section II. For this reason the "t" was calculated to determine if oral administration of this test could account for a significant rise in mean score. No significant difference was found.[1] In view of this it seems that the reliability of the VICS in Section III is lower not because of the variable of methods of administration but instead, it may be speculated, because of the use of a different group of subjects.[2]

On the other hand, there are certain similarities in results obtained in both Sections II and III. For instance, Vocabulary and Information seem to be consistently the most stable sub-tests. The coefficients obtained on the Comprehension sub-test were almost similar in both cases. Similarities sub-test was consistently lowered in reliability. It is interesting to note that although the Information sub-test rated as one of the most reliable sub-tests in this self-administering form, the mean score was lowered over time in both

---

1. A "t" of .522 was obtained which was not statistically significant.

2. Two different groups were used in Section II and Section III. Group I was the more homogeneous group (revealing a narrow S.D.). This could account for partial lowering of the results in Section III.

analysis of Section II and Section III.

In analyzing the accuracy of prediction of the test employed in this study, it was found that the percentages of errors accounted for in predicting R.N. scores by the VICS, A.C.E. and Bellevue Full Scale were 2%, 7.5% and 8.7% respectively. Thus, it appears that if one wishes to consider the selection of nurses for training at the University of Alberta Hospital on the basis of the VICS, then this test would be more efficient in accounting for 2 unsuccessful students out of 100 who have applied (i.e. these 2 students would not pass the R.N. examinations) than if the VICS had not been employed at all as a selection device.[1] On the other hand, the A.C.E. and the Bellevue Full Scale seem to surpass the VICS in reducing errors in prediction. (However, it should be noted that for a more heterogeneous group of student nurses, these percentages may increase somewhat).

---

1. If the reader is interested in employing the VICS for selection, the merit of its inclusion would appear to depend upon the particular practical circumstances in the hospital (e.g. economic expenditures per student, dormitory facilities, etc.). In this case further analysis of the predictive qualities of this test may be suggested and can be readily calculated from the correlation coefficients already obtained in this study. Further interpretation regarding the predictive value of the VICS on this group may be gained from the computing of the Coefficient of Alienation or the Coefficient of Determination. In particular, the use of cutting scores is suggested (ibid - p.32, ref. footnote). However, for purposes of this study these analyses were not attempted in view of the results already obtained.

Discussion

Since the reliability coefficients as well as all the validity coefficients are low, it appears that there may be a variable factor throughout which has had an influence on these obtained correlations and which has had the effect of lowering the results in this investigation.

Thus, calculations were done for restriction of range between the curtailed VICS and the Bellevue Full Scale.[1] The resulting correction was a correlation coefficient of .74 as contrasted with the .42 obtained in this investigation. It is evident from this that the homogeneity of the sample selected for this study has been one factor in making the coefficients "spuriously" low.[2] It is also possible that other variable errors of measurement exist. However, the correction for attenuation was not attempted. Sufficient data were not available for this analysis.[3]

Despite the fact that the correlation coefficients obtained

---

1. The standard deviation of 19.8 for the mean age of 18.5 of the Wechsler-Bellevue Full Scale was employed in this computation. Wechsler (29), 3rd Ed., Table 14, p. 118.

2. Further analysis using the Coefficient of Determination (ibid - p. 32, ref. footnote) reveals that the restriction of range accounts for approximately 17% of the variable affecting the validity coefficients.

3. In order to use the Spearman-Brown correction effectively, the reliability coefficients of one of the other two tests of intelligence or the criterion should be obtained. McNemar (29). It is difficult to find reliability coefficients of a sample representative of the group used in this project on the A.C.E. or W.B.F.S. In addition, there is no available data on the R.N. reliabilities.

may be assumed to be higher in view of the restriction in range of the sample, the results do not support the major hypothesis as stated. In fact, the reverse seems to be true. The A.C.E. and particularly the Bellevue Full Scale reveal better correlations with the criterion of success than the VICS. Some factors which might have had an influence on this reversal were analyzed. No definite indication was found from this analysis that the type of sub-tests (Verbal or Performance) or individual sub-tests selected in the self-administering form had a bearing on lowering the validity of the VICS with the criterion of success. Generally, in view of the non-significant lowered positive correlations of the VICS with the R.N. marks and the lowered Index of Forecasting Efficiency, the advisability of the use of this test for prediction of R.N. scores on a group with such select entrance requirements is questionable. Although the increase in accuracy of prediction by employing the VICS appears to be low on this particular group, it may have some practical value as part of a battery for selection of nurses in training, particularly on a more heterogeneous group of student nurses.

There is an indication from the results that this shortened form seems to be a more valid indicator for measuring the factors tested by the A.C.E. and the Wechsler-Bellevue Full Scale. In reference to the Bellevue Scale, the results obtained are somewhat in accordance with the findings from other short-form studies where good correlations were obtained with the Full Scale. The lowered correlation obtained in this study because of the homogeneous group employed seems to substantiate McNemar's (30) criticism of the other

short-form studies.[1] Relatively, though, the VICS compares favorably with other short-form tests of the Bellevue.

The calculations discerning the relationship between the VICS and the A.C.E. reveal some interesting results. The degree of agreement in measurement between these two tests is high and both reveal lower correlations with the R.N. marks than does the Bellevue Full Scale. It appears that the VICS may be measuring many of those aspects of academic achievement with which the A.C.E. is concerned,[2] and which may not be too directly related to success measured by the R.N. marks.

The reliability of this short-form seems to be established from the findings of this investigation. It appears that the VICS can be considered a reliable measure over time. This test seems to be more reliable when the administration conditions are constant as in Section II. The variable of the examiner administering the test seems to have a slight effect in lowering the reliability. However, since no significant difference was found between the means on retest, it is possible that this conclusion may be only tenuous. The lowered reliability coefficient in Section II may also be a function of the use of a different sample rather than a change in administration conditions.

More fluctuation in measurement was noted between the individual sub-tests. Vocabulary and Information revealed consistently

---

1. p. 13

2. p. 30 f.

higher correlations. The use of the Information sub-test is questionable for inclusion in this group, other than as a negative indicator. The mean score was lowered on this test over time which might suggest the consideration of possible extinction of the abilities measured by this test during nursing training. The inter-test variability among the individual sub-tests does not seem to differ too markedly for this group than from other reliability studies conducted by Derner and Gilhooly on different groups.

Finally, it should be noted that all correlations obtained in this study were low and indications derived from the results have been considered on the basis of the .05 significance level. Caution is advised in interpreting these results which have been found to be "spuriously" low, in part, to a restricted range of sample. Other statistical tests for variables which might have further affected the results obtained were not attempted. In view of this, conclusions may be considered only as tenuous.

Chapter V

SUMMARY AND CONCLUSIONS

This investigation was concerned with a check on the validity of a Verbal Self-Administering Short-Form of the Wechsler-Bellevue Intelligence Scale on a group of student nurses. An hypothesis was made that "a Verbal Self-Administering Short-Form of the Wechsler-Bellevue Adult Intelligence Scale was a more accurate predictor of success than the A.C.E. or the Wechsler-Bellevue Full Scale on a group of student nurses".

An assembly of the Vocabulary, Information, Comprehension and Similarities sub-tests was adapted for purposes of study. To enable a further shortening of administration time, a self-administering form of these sub-tests was developed.

The sample consisted of an all female group of student nurses, who were registered in the Diploma Course in Nursing at the University of Alberta Hospital. Originally, the entire number of two consecutive classes was used for study. However, during the course of training only 61 subjects remained for study. Thirty-one of this group constituted Group I of this investigation and thirty were in Group II. These groups were matched on the basis of High School averages and the "t" critical ratio was employed to determine the significance of difference between the means. No significant difference was found and the groups were considered comparable. The age range of the Total Group was 18-22 years. The mean age for the Total Group, Group I and Group II was as follows: 18.67, 18.87 and 18.33. The mean High School averages

and the standard deviations for the Total Group, Group I and Group II were 65.63, 6.94; 65.65, 6.40; and 66.41, 7.47 respectively.

The procedure consisted of administering the VICS and the A.C.E. to the Total Group in 1953-54. Two years later the Wechsler-Bellevue Full Scale was administered to Group I. Group II was re-tested on the VICS at this time to facilitate a test-retest reliability study on this Verbal Self-Administering Short-Form. The criterion for success was the Registered Nurses standard score, which was obtained one year later, after class graduation.

To test the major hypothesis that the VICS predicts success better than the A.C.E. or the Wechsler-Bellevue Full Scale, the Pearson Product Moment correlation coefficients were computed between the VICS, the A.C.E., the Wechsler-Bellevue Full Scale and the criterion of success, the Registered Nurses marks. An additional check on validity was concerned with comparing the short-form with other tests of intelligence. Correlations were calculated between the VICS, A.C.E. and Full Scale for this purpose. In all cases where Full Scale data was used the analysis was conducted on the smaller Group I.

The reliability of this Self-Administering Short-form test was also determined. Two analyses were made in this instance. The first included a test-retest of this Self-Administering form on Group II. For interest, the second analysis included the measurement of the effect of the variable of method of administration by correlating the VICS self-administering form and the VICS examiner administered test of a test-retest study on Group I.

In summary, all correlations were found to be positive but

low. It was determined that these lowered correlations were, in part, due to the homogeneous sample used in this investigation. Tentative interpretations only have been made on the basis of the significance at or beyond .05. These were:

1. The VICS does not predict R.N. success better than the A.C.E. or Wechsler-Bellevue Full Scale. In fact, the results obtained between the VICS and the criterion of success were far below those of the other two longer tests. The use of this short form test as a single device for predicting Registered Nurses' marks on this sample does not seem to be warranted.

2. The VICS revealed a better degree of agreement of measurement with the Wechsler-Bellevue Full Scale and particularly the A.C.E. This indicates the possibility of the use of this test as a substitute for rough screening purposes wherever these longer tests may be employed. The added advantage of the VICS in relation to its use instead of the Bellevue Scale is that this shorter form of test can be administered in group form. In considering the A.C.E. the advantage of the VICS is that it still retains some of the qualitative clinical features of the Full Scale.

3. The VICS self-administering form is a reliable test over time. It appears that changes in method of administration from a self-administering form to an examiner administered form lowered the reliability somewhat. However, it was questionable whether the lowered coefficients obtained when the examiner administered

form was employed were due to less objectivity in administration or due to a function of a different group's performance. It may be tentatively concluded that the method of administration has little, if any, effect on the reliability of this test.

4. Intra-test variability of the four sub-tests is similar in degree of fluctuation to other findings reported in the literature. The Information and especially Vocabulary revealed consistently higher reliability coefficients, whereas a greater degree of fluctuation was found on the Similarities and Comprehension sub-tests.

5. Although the Information sub-test was a reliable measure on this group, the effect of this test over a period of training is negatively influenced.

Chapter VI

RECOMMENDATIONS

From the findings of this investigation, the VICS test for prediction of R.N. marks on the student nurses at the University of Alberta Hospital does not seem to be particularly advantageous. Because of the rigid, precise methods of selection and the resulting homogeneity of the sample of nurses in training at the University of Alberta Hospital, further research concerning the use of this short-form for this specific group does not appear to be warranted. In the area of the effect of examiner administration vs self-administration of this test, more extensive study concerned entirely with determining the effects of these differences may be recommended. For it should be noted in the present study that in order to facilitate a study of validity and reliability, adequate counter-balancing of the sample in the reliability section was curtailed. Consequently, it was difficult to determine whether the differences in the reliability coefficients obtained were due to the effect of change in method of administration or purely to a function of different subjects' performance.

Although this test does not appear to be a good predictor of R.N. marks, there is the possibility that the VICS may be a useful measure of predicting academic success in the areas where the matriculation records are not as complete. Some suggestions for future investigations with this short-form should be considered. There may be some merit for its use in predicting academic success on a more heterogeneous sample of student nurses, where pre-selection on the

basis of High School averages is not uniform. Such a short time-saving device as the VICS, which is also clinically adaptable, may be useful as a rough screening device for selecting applicants for nurses training. It may also be suggested that the VICS could be considered for prediction of academic success in other areas as distinct from determining those abilities necessary for a successful nurse. Specifically, the use of the VICS as a predictor of success at the college level may be suggested. There are several indications from this investigation for further study in this area in addition to the findings of Merril & Anderson. For instance, the student nurses sample selected may not be representative of college groups. Also, there may be some different factors in the measurement of college achievement which are not directly related to the R.N. examinations. Some support for this can be derived from the higher correlations of the VICS with the Bellevue Full Scale or the A.C.E. obtained in this study.

Finally, since this study has been one of the first attempts reported in validation of a short-form test for prediction of academic success, it may be that the entire area of developing a shorter-form for prediction should be seriously reconsidered. When one is concerned with future prediction many facets of behavior are involved. It follows then that there may be a danger in selecting one test which may measure only a small segment of this ability to succeed. This would be particularly noticeable when the quantitative aspects of measurement are further reduced as by a short-form test. Thus it may not be surprising that this short-form or, for that matter any short-form which is correlated with a criterion of academic success, would reveal low results

regardless of the type of subjects involved. The advisability of using short-forms for prediction of scholastic achievement may be questioned. Caution is suggested in considering further efforts in this direction.

BIBLIOGRAPHY

1. Anderson, E.C. et al. "A Comparison of the Wechsler-Bellevue, Revised Stanford Binet and American Council on Education Tests at the College Level", J. Psychol., 14:217-26, 1942.

2. Armitage, S.G. and Pearl, D. "Wechsler-Bellevue Changes Over Time", J. Clin. Psychol., 14:22-4, 1958.

3. Bay, M. and Berks, . "Comprehension, Similarities and Digit Symbols on the Wechsler-Bellevue Scale used as a Court Device", American Psychologist, 3:36, 1948, (Psychological Abstract, 743, 1949).

4. Brown, M. and Bryan, G.F. "Sex Differences in Intelligence", J. Clin. Psychol., 11:303-4, 1955.

5. Campbell, W.G. Form and Style in Thesis Writing, Boston: Houghton Mifflin Co., 1954.

6. Cole, A.H. and Bigelow, K.W. A Manual of Thesis Writing for Graduates and Undergraduates, New York: John Wiley and Sons, Inc., 1935.

7. Cohen, B. "Is there Examiner Bias on the Wechsler-Bellevue?" Proc. Okla Acad. Sc., 31:150-3, 1950, (Psychological Abstract, 417, 1953).

8. Cohen, B. and Stiesel, I. "An Evaluation of the Various Short Forms of the Wechsler-Bellevue, used as a Court Device", American Psychologist, 3:30, 1948, (Psychological Abstract, 743, 1949).

9. Cotzin, M. and Gallager, J.J. "Validity of Short-Forms of the Wechsler-Bellevue for Mental Defectives", J. Consul. Psychol., 13:357-65, 1949.

10. Cotzin, M. and Gallager, J.J. "The Southbury Scale: A Valid Abbreviated Wechsler-Bellevue for Mental Defectives", J. Consul. Psychol., 14:358-64, 1950.

11. Cummings, S.B., MacPhee, H.M. and Wright, H.F. "A Rapid Method of Estimating the I.Q. of Sub-normal White Adults", J. Psychol., 21, 1946.

12. Derner, G.F. and Aborn, M. "The Reliability of the Wechsler-Bellevue Sub-tests and Scales", J. Consul. Psychol., 14:172-9, 1950.

13. Finkelstein, M., Gerboth, R. et al. "Standardization of a Short-Form of the Wechsler Vocabulary Sub-test", J. Clin. Psychol., 133-5, 1952.

14. Garrett, H.R. Statistics in Psychology and Education, Toronto: Longhmans, Green and Co., 3rd Edition, 1947.

15. Geil, G.A. "A Clinically Useful Abbreviation of the Wechsler-Bellevue Test", J. Psychol., 20:101-8, 1945.

16. Gerboth, R. "A Study of Two Forms of the Wechsler-Bellevue Intelligence Scales", J. Consul. Psychol., 14:365-70, 1950.

17. Gilhooly, F. "Wechsler-Bellevue Reliability and the Validity of Certain Diagnostic Signs of the Neuroses", J. Consul. Psychol., 14:82-7, 1950.

18. Gurvitz, M.S. "An Alternate Short-Form of the Wechsler-Bellevue", American J. Orthopsychiatry, 15:727-33, 1945.

19. Gurvitz, M.S. "Some Defects of the Wechsler-Bellevue Scale", J. Consul. Psychol., 16:124-6, 1952.

20. Hamister, R.C. "The Test-Retest Reliability of the Wechsler-Bellevue Intelligence Test (Form I) for a Neuropsychiatric Population", J. Consul. Psychol., 13:39-43, 1949.

21. Herring, F.H. "The Evaluation of Published Short-Forms of the W.B. Scale", J. Consul. Psychol., 16:119-23, 1952.

22. Hilden, A. et al. "Empirical Evaluation of Short Wechsler-Bellevue Scales", J. Clin. Psychol., 8:323-31, 1952.

23. Hunt, W.A. and Stevenson, . "Psychological Testing in Military Clinical Psychology", Psychol. Review, 53:25-35, 1946.

24. Hunt, W.A. et al. "The Validity of Some Abbreviated Individual Intelligence Scales", J. Consul. Psychol., 12:18-52, 1948.

25. Hunt, W.A. et al. "Further Standardization of the C, V, S, Individual Intelligence Scale", J. Consul. Psychol., 12:355, 1948.

26. Hunt, W.A. et al. "Clinical Possibilities of an Abbreviated Individual Intelligence Scale", J. Clin. Psychol., 3:209-16, 1947.

27. Kriegman, G. and Hansen, F. "VIBS - A Short-Form of the Wechsler-Bellevue Intelligence Scale", J. Clin. Psychol., 3:209-16, 1947.

28. Lindquist, E.F. A First Course in Statistics, Cambridge, Mass: Houghton Mifflin Co., Revised edition, 1942.

29. McNemar, Q. Psychological Statistics, New York: John Wiley and Sons, Inc., 2nd edition, 1955.

30. McNemar, Q. "On Abbreviated W.B. Scales", J. Consul. Psychol., 14:79-81, 1950.

31. Merril, R. and Heathing, L. "Centile Scores for the W.B. on a University Counselling Centre Group", J. Consul. Psychol., 16:406-9, 1952.

32. Merril, R. "A Comparison of the Wechsler-Bellevue and A.C.E. Tests on University Counselling Center Group", J. Consul. Psychol., 17:63-6, 1953.

33. Murphy, D.B. and Langston, R.D. "A Short Form of the Wechsler-Bellevue and the Army Classification Battery as a Measure of Intelligence", J. Consul. Psychol., 20:405, 1956.

34. Norman, B. "Sex Differences and other Aspects of Young Superior Adults on the W.B.", J. Consul. Psychol., 17:417-18, 1953.

35. Patterson, C.H. "A Comparison of Various 'Short Forms' of the Wechsler-Bellevue Scale", J. Consul. Psychol., 10:260-67, 1946.

36. Patterson, C.H. Counselors Manual. 14-25, 98-113.

37. Patterson, C.H. "A Further Study of Two Short-Forms of the Wechsler-Bellevue Scale", J. Consul. Psychol., 12:147-58, 1948.

38. Rabin, A.I. "A Short-Form of the Wechsler-Bellevue Test", J. Applied Psychol., 27:320-24, 1943.

39. Rabin, A.I. and Guertin, . "Research with the W.B. Test", Psychol. Bull., 48:211-16, 1951.

40. Rabin, A.I. "Use of the Wechsler-Bellevue Scale with Normal and Abnormal Persons", Psychol. Bull., 42:410-22.

41. Rapaport, D., Geil, M. and Schafer, R. Diagnostic Psychological Testing, Vol. 1, Chicago: The Year Book Publishers, Inc., 1945.

42. Sartain, A.Q. "A Comparison of the new Revised Stanford Binet, the Bellevue Scale and Certain Group Tests of Intelligence", J. Soc. Psychol., 23:237-9, 1946.

43. Siegel, S. Non-Parametric Statistics, New York: McGraw-Hill Book Co., 1956.

44. Springer, N.W. "A Short-Form of the Wechsler-Bellevue Test as Applied to Naval Personnel", American J. Orthopsychiatry, 16:341-3, 1946.

45. Sloan, W. "The Development of a Wechsler-Bellevue II Short Form", Pers. Psychol., 8:347-53, 1955, (Psychological Abstract, 7838, 1956).

46. Thurstone, L.L. and Thurstone, T.G. Manual of Instructions of the American Council on Education Psychological Examination, Princeton, New Jersey: Co-operative Test Division, Educational Testing Service, 1950.

47. Watson, A. Contributions Towards Medical Psychology, Theory and Diagnostic Methods, Volume II, New York: The Ronald Press Co.

48. Webb, W.B. and Dehaan, H. "Wechsler-Bellevue Split-Half Reliabilities in Normals and Schizophrenics", J. Consul. Psychol., 15:68-71, 1951.

49. Wechsler, D.H. Measurement of Adult Intelligence, Baltimore: Williams and Wilkins Co., 3rd edition, 1944.

50. Wechsler, D.H. The Measurement & Appraisal of Adult Intelligence, 4th edition.

APPENDIX "A"

VERBAL INFORMATION TEST (VICS)

Date _____

Name _____ Age _____

Address _____ Phone _____

Education _____

Occupation _____

INSTRUCTIONS:

Please answer the questions on the following pages. Enter your responses in the blank spaced provided. Write as clearly as possible. Work as quickly as you can.

VICS 3 - 7/55/-150

SECTION I

Answer these questions:

1. Who is the Prime Minister of Canada?

   Who was Prime Minister before him?

2. What is a thermometer?

3. What does rubber come from?

4. Where is London?

5. How many pints make a quart?

6. How many weeks are there in a year?

7. What is the capital of Italy?

8. What is the capital of Japan?

9. How tall is the average Canadian woman?

10. Who invented the aeroplane?

11. Where is Brazil?

12. How far is it from Paris to New York?

13. What does the heart do?

14. Who wrote Hamlet?

15. What is the population of the United States?

16. When is Queen Victoria's birthday?

17. Who discovered the North Pole?

18. Where is Egypt?

19. Who wrote Huckleberry Finn?

20. What is the Vatican?

21. What is the Koran?

22. Who wrote Faust?

(continued on next page)

Section I - continued

23. What is an Habeas Corpus?

24. What is Ethnology?

25. What is the Apocrypha?

(Go on to the next page)

SECTION II

Answer these questions:

1. What is the thing to do if you find an envelope in the street, that is sealed, and addressed and has a new stamp?

2. What should you do if while sitting in the movies (theatre) you were the first person to discover a fire (or see smoke and fire)?

3. Why should we keep away from bad company?

4. Why should people pay taxes?

5. Why are shoes made of leather?

6. Why does land in the city cost more than land in the country?

7. If you were lost in a forest (woods) in the daytime, how would you go about finding your way out?

8. Why are laws necessary?

9. Why does the state require people to get a license in order to be married?

10. Why are people who are born deaf usually unable to talk?

(Go on to the next page)

SECTION III

In the list below you will find two things which are the same or alike in certain ways. You are requested to write down in what way they are alike.

For example: "In what way are an orange and a banana the same?"

For this you might say: "You can eat them both" or "They both have skins" or "They are both fruits".

They are alike in these ways. Write down ONE way in which they are alike.

1. In what way are an orange and a banana the same?
2. In what way are a coat and a dress the same?
3. In what way are a dog and a lion the same?
4. In what way are a wagon and a bicycle the same?
5. In what way are a daily paper and a radio the same?
6. In what way are air and water the same?
7. In what way are wood and alcohol the same?
8. In what way are eye and ear the same?
9. In what way are egg and seed the same?
10. In what way are poem and statue the same?
11. In what way are praise and punishment the same?
12. In what way are a fly and a tree the same?

(Go on to the next page)

SECTION IV

Give the meaning of each of the following words:

1. apple
2. donkey
3. join
4. diamond
5. nuisance
6. fur
7. cushion
8. shilling
9. gamble
10. bacon
11. nail
12. cedar
13. tint
14. armory
15. fable
16. brim
17. guillotine
18. plural
19. seclude
20. nitroglycerine
21. stanza

(continued on next page)

Section IV - continued

22. microscope
23. vesper
24. belfry
25. recede
26. affliction
27. pewter
28. ballast
29. catacomb
30. spangle
31. espionage
32. imminent
33. mantis
34. hari-kiri
35. chattel
36. dilatory
37. amanuensis
38. proselyte
39. moiety
40. aseptic
41. flout
42. traduce

APPENDIX "B(i)"

Department of Psychology

University of Alberta

Dear Miss _____:

You will remember, early in October, I discussed with your class the intention of re-testing your group on an individual basis. Arrangements have now been made for the testing to begin on Monday, November 28th, and administration will continue until testing has been completed.

For research purposes the class has been divided into two groups. Those comprising Group I will be tested first. Group II will be tested at a later date. The results will be considered in group form and will be of a statistical nature. This will protect your identity in the project. May I again mention the importance of not discussing the test items among yourselves since it may influence the results as a whole.

Miss Swanson, your class president has a list of Group I. If you find that you are unable to attend at the specified time, it would be appreciated if you would check with her and arrange a substitute for your appointment.

An appointment has been made for you on _____ at _____ in classroom 2 of the Nurses' residence. The test time will be approximately one hour and one half, or less.

I would like to thank you again for your co-operation and willingness to participate in this study which, it is hoped, will be a valuable contribution to psychological research.

Sincerely,

REA/sg

Rita E. Aldridge

APPENDIX "B(ii)"

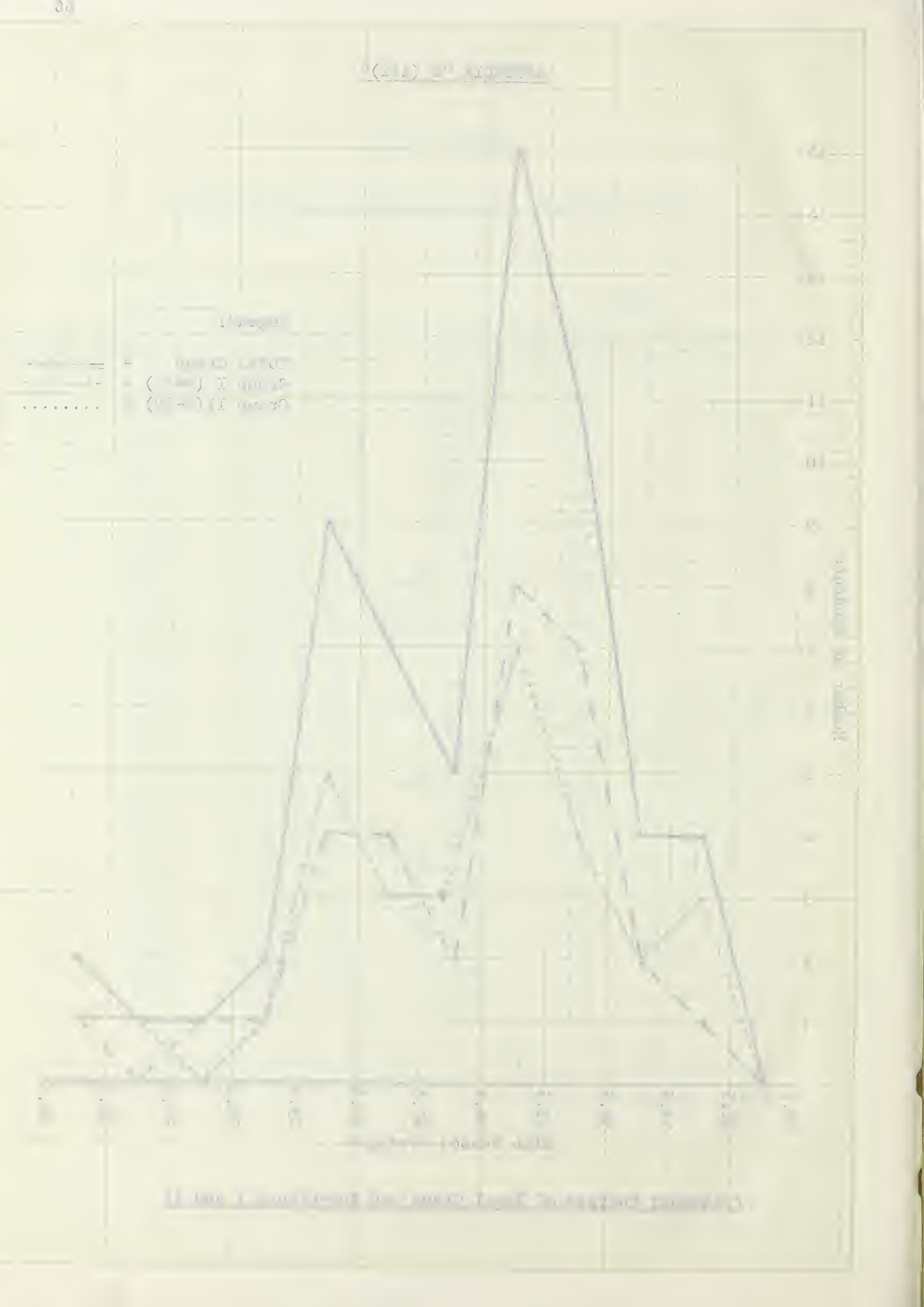
Data of Sample: Age, H.S. Averages and Shift Work

Group I				Group II			
N	Age	H.S.Average	Shift	N	Age	H.S.Average	Shift
1	18	61.2	Night	1	19	62.1	Classes
2	18	70.5	Afternoon	2	18	73.7	"
3	18	69.4	Night	3	19	62.3	"
4	20	60.0	Night	4	18	85.25*	"
5	19	55.7	Night	5	18	62.3	"
6	18	65.6	Day Off	6	18	60.1	"
7	18	77.1	Night	7	18	53.6	"
8	19	68.0	Night	8	18	68.4	"
9	19	60.7	Night	9	19	67.4	"
10	18	64.0	Afternoon	10	19	55.85*	"
11	18	59.9	Afternoon	11	19	59.9	"
12	18	63.5	Split shift	12	19	67.1	"
13	18	61.1	Afternoon	13	19	58.7	"
14	18	56.1	Night	14	19	73.0	"
15	20	72.3	Night	15	18	63.7	"
16	18	83.25*	Afternoon	16	18	72.2	"
17	22	60.4	Night	17	18	62.0	"
18	18	56.1	Night	18	18	72.1	"
19	18	64.3	Day Off	19	18	59.3	"
20	20	70.6	Afternoon	20	18	54.1	"
21	19	62.0	Night	21	19	61.1*	"
22	20	65.4	Night	22	19	71.6	"
23	19	75.0	Afternoon	23	18	62.1	"
24	18	71.4	Afternoon	24	18	66.8	"
25	19	62.4	Afternoon	25	18	68.4	"
26	22	64.1	Afternoon	26	18	63.3	"
27	19	72.5*	Afternoon	27	18	80.4*	"
28	20	72.9	Night	28	18	76.5*	"
29	18	61.7*	Day Off	29	18	68.0	"
30	19	64.6	Night	30	18	57.1	"
31	19	63.6	Night				

* Out-of-province averages

APPENDIX "B (iii)"

Frequency Polygon of Total Group and Sub-Groups I and II



APPENDIX "C"

Standard scores are given on the five subjects which constitute the Registered Nurses examinations. These are (1) Medical Nursing, (2) Surgical Nursing, (3) Pediatrics, (4) Obstetrics and (5) Psychiatric Nursing. These scores are calculated on the basis of all candidates in all jurisdictions and are comparable from subject to subject, year to year and school to school. (The mean standard score is 500).

The R.N. examinations are copywrited and, for this reason, copies of previous examinations were not available for this present study. However, with the co-operation of Miss Clark, Director of Nursing for the University of Alberta Hospital, some idea of the type of examination items was obtained. The R.N. examinations are designed for professional selection of "safe" nurses, i.e. the theoretical knowledge as well as practical application is stressed. The items on this examination are presumably "situational" type items, where the student is required to select the best answer from four possible given procedures.

APPENDIX "D"

(i) Inter-Correlations of VICS, R.N. Marks & A.C.E. on Group I (N=31)

		r
VICS vs R.N.	=	.16
A.C.E. vs R.N.	=	.33
VICS vs A.C.E.	=	.40

(ii) Comparison of Verbal & Performance Scales with the Criterion (N=31)

		r
Verbal vs R.N.	=	.39
Performance vs R.N.	=	.26

(iii) V, I, C & S (VICS 1953-54 Administration) Correlated with R.N. (N=31)

		r
V vs R.N.	=	.32
I vs R.N.	=	.08
C vs R.N.	=	.26
S vs R.N.	=	.13

Note: The self-administering form (1955) was used in this analysis. In view of the lowered results, combinations of these four tests were not attempted.